行政院國家科學委員會專題研究計劃成果報告

當變異數不等時單階段常態分配與平均的多重比較程序

Single Stage Multiple Comparison Procedures with the Average for Normal Means

計劃編號: NSC88-2118-M-032-007

執行期限: 87 年 8 月 1 日 至 88 年 7 月 31 日

主持人: 吳淑妃 淡江大學統計系

一. 中文摘要

在這篇報告裡, 我們提出了單階段抽樣程序來解決當變異數不等時常態分配平均數與平均的多重比較程序的問題. 我們的多重比較程序主要分成兩方面來討論, 一為子集選擇程序, 另為等尾同時信賴區間. 這些程序應可廣氾的應用在選出比平均好的常態分配平均數上,或是把 k 個常態分配平均數分成比平均好,比平均差,及和平均沒什麼差異的三個群體.當樣本大小相等時, 我們用 Monte-Carlo 方法, 找出模擬臨介值的表供使用者使用.

關鍵辭:

子集選擇程序,同時信賴區間,Bonferroni

ABSTRACT

In this article, multiple comparison procedures with the average for normal means under heterocedasity are under investigation by single stage sampling technique. A subset selection approach and a simultaneous confidence interval (S.C.I) approach are considered for normal distribution. These procedures will have broad applicability in selecting a subset which includes all populations with larger means than the average in experimental designs and/or in identifying groups of treatments with smaller than the average, larger than the average and not much difference from the average means in various fields. A simulation program by Monte-Carlo method is developed to find the simulated critical values for both of subset selection and SCI. The simulated values are tabulated for practice use for equal sample size.
Key Words and Phrases:
subset selection, simultaneous confidence interval, Bonferroni inequality,
Monte Carlo technique.

1. 計劃緣由與目的

When working with statistical data analysis one often has only one single sample available. Chen and Lam (1989) proposed a single-sample procedure by splitting up the one sample into two portions for the problem of interval estimation. Lam (1992) adapted it to a subset selection procedure which was proven to be quite satisfactory. Wen and Chen (1994) applied the single-sample procedure for multiple comparisons with the largest mean and with a control. The single-sample procedure has design and computational simplicity in practice. The single-stage sampling procedure for multiple comparisons with the average of all populations under consideration is proposed below.

Let X_{ij} be an independent random sample of size $n_i \geq 3$ from the normal population π_i with unknown mean μ_i and unknown and unequal variance σ_i^2. Employ the first (or randomly) n_i-1 observations to define the usual sample mean and sample variance, respectively, by

$\bar{X}_i = \sum_{j=1}^{n_i-1} X_{ij} / n_i - 1$ and

$S_i^2 = \sum_{j=1}^{n_i-1} (X_{ij} - \bar{X}_i)^2 / (n_i - 2)$.

Let the weights of the observations be

$$U_i = \frac{1}{n_i} + \frac{1}{n_i}\sqrt{\frac{1}{n_i-1}[S_{[k]}^2 / S_i^2 - 1]},$$

$$V_i = \frac{1}{n_i} - \frac{1}{n_i}\sqrt{(n_i-1)[S_{[k]}^2 / S_i^2 - 1]},$$

where $S_{[k]}^2$ is the maximum of $S_1^2, \ldots, S_k^2$.

Let the final weighted sample mean be defined by $\tilde{X}_i = \sum_{j=1}^{n} W_{ij} X_{ij}$, where

$$W_{ij} = \begin{cases} U_i & for \quad 1 \leq j \leq n_i - 1 \\ V_i & for \quad j = n_i \end{cases}$$

二. 結果與討論

As for the problem of multiple comparison procedure with the average normal mean under heteroscedasticity, the set of upper one-sided confidence intervals is given

by $\tilde{U}_i = (-\infty, \tilde{X}_i - \bar{\tilde{X}} + d^* S_{[k]} / \sqrt{n_i}), i = 1, \ldots, k$

where $d^* > 0$, $\upsilon_i = n_i - 2$, and

$\bar{\tilde{X}} = \sum_{i=1}^{k} \tilde{X}_i / k$ d Then the probability of inclusion of $\theta_i - \bar{\theta}$ in $\tilde{U}_i$ is given by

$$P(\theta_i - \bar{\theta} \in \tilde{U}_i, i = 1, \ldots k)$$

$$= P(\frac{k-1}{k}\tilde{Y}_i - \sum_{j \neq i}^{k} \tilde{Y}_j / k > -d^* S_{[k]} / \sqrt{n_i}, i = 1, \ldots, k)$$

$$= P(\frac{k-1}{k} T_i - \sum_{j \neq i}^{k} \sqrt{n_i / n_j} T_j / k > -d^*, i = 1, \ldots, k)$$

$$= P(\tilde{T}_i < d^*, i = 1, \ldots, k)$$

where $\tilde{Y}_i = \tilde{X}_i - \theta_i$, $T_i = \sqrt{n_i}\tilde{Y}_i / S_{[k]}$ has t distribution with υ_i df, and

$$\tilde{T}_i = \frac{k-1}{k} T_i - \sum_{j \neq i}^{k} \sqrt{n_i / n_j} T_j / k, i = 1, \ldots, k$$

It is difficult to find the exact joint sampling distribution of the singular k-variate statistic of $\tilde{T}_1, \ldots, \tilde{T}_k$. Therefore, the Monte-Carlo simulation is used here to obtain an approximate sampling distribution of the maximum order statistic of $\tilde{T}_{[k]}$. In order to have the probability of cinclusion being at least P^*, the value of d^* is given by the P^* \$th percentile of the approximate sampling distribution of $\tilde{T}_{[k]}$, $P(\tilde{T}_{[k]} < d^*) = P^*$.

Likewise, the $(1-\alpha)100\%$ SCI of $\theta_i - \bar{\theta}$ is given by

$$(\tilde{X}_i - \bar{\tilde{X}} - \tilde{d} S_{[k]} / \sqrt{n_i}, \tilde{X}_i - \bar{\tilde{X}} + \tilde{d} S_{[k]} / \sqrt{n_i}), i = 1, \ldots, k$$

where $\tilde{d}$ is given by the upper $\alpha / 2$ percentile of the approximate sampling distribution of $\tilde{T}_{[k]}$, $P(\tilde{T}_{[k]} < \tilde{d}) = 1 - \alpha / 2$.

In the case of equal sample size $n_i = n$, the simulated value of d^* for subset selection when $P^* = .95$ and the simulated value of $\tilde{d}$ for simultaneous confidence interval when $1-\alpha = .90$ are given in the following table. The upper entry is the average of the simulated percentiles, each based on 3000 runs, and the standard error (s.e.) is reported at the lower entry. To do this we only need to generate some random numbers from a standard normal or chi-square distribution with some specific degrees of freedom and calcualte $\tilde{T}_i$'s, then order these $\tilde{T}_i$'s and select the maximum value $\tilde{T}_{[k]}$ at each run. After N independent runs, the P^*th percentile of the distribution of $\tilde{T}_{[k]}$ can be approximated by selecting the NP^* biggest number out of the ranking list of the N largest values of $\tilde{T}_{[k]}$.

n	[illegible]	[illegible]	[illegible]	5	[illegible]	[illegible]	[illegible]	9	10
[illegible]	1[illegible]	.19.	1.	[illegible]	.9	.5	51.	.5	.0
	0.0	0.05	0.0	0.0	0.0	0.11	0.1	0.1	0.1
[illegible]	[illegible]	.51	5.	.1	.0	[illegible]	.0	[illegible]	.1
	0.0	0.05	0.0	0.05	0.0	0.05	0.05	0.0	0.05
5	[illegible]	.0	.51	.91	[illegible]	.5	[illegible]	.11	5.
	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.05
[illegible]	1.9	.5	.95	[illegible]	.9	.5	.9	.0	.1
	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
[illegible]	1.	[illegible]	[illegible]	.9	.11	[illegible]	[illegible]	.0	.0
	0.0	0.01	0.01	0.0	0.01	0.0	0.0	0.0	0.0
[illegible]	1.	.0	.51	.9	.09	.1	[illegible]	[illegible]	[illegible]
	0.01	0.01	0.01	0.0	0.0	0.01	0.0	0.0	0.0
9	1.	.11	.0	.1	.9	.0	.1	[illegible]	.1
	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.0
10	1.	.0	.3	.5	[illegible]	[illegible]	.95	.0	.09
	0.01	0.01	0.01	0.01	0.01	0.0	0.0	0.0	0.01
15	1.5	1.9	.1	[illegible]	[illegible]	.5	[illegible]	.5	[illegible]
	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
[illegible]0	1.	1.	.10	[illegible]	.0	[illegible]	.5	[illegible]	.0
	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
[illegible]0	1.5	1.	.0	.19	.1	.9	.5	.55	.0
	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01

四. 計劃結果與自評

我已完成理論的推導以及模擬的研究，所有的程式皆使用 C++語言. 此研究結果可有效的解決當變異數不等時常態分配與平均的多重比較程序的問題，相信很快就可在學術學刊發表.

五. 參考資料

Chen, H. J. and Dudewicz, E. J. (1976). Procedures for fixed-width interval estimation of the largest normal means. *Journal of*

American Statistical Association, 71, 752-756.

Chen, H. J. and Lam, K. (1989). Single-stage interval estimation of the largest normal mean under heteroscedasticity. Communication in Statistics-Theiry and Methods, 18(10), 3703-3718.

Dudewicz, E. J. and Dalal, S. R. (1975). Allocation of observations in ranking and selection with unequal variances. *Sankhya, Ser*. B., 37, 28-78.

Gupta, S. S. and Panchapakesan, S. (1979). *Multiple Decision Procedure: Theory and Methodology of Selecting and Ranking Populations*. Wiley, New York.

Hochberg, Y. and Tamhane, A. C. (1987). *Multiple Comparison Procedures*. Wiley, New York.

Lam, K. (1992). Subset selection of normal populations under heteroscedasticity. The Frontiers of Modern Statistical Inference Procedures, II (Eds. Bofinger, E., Dudewicz, E. J., Lewis, G. J., and Mengersen, K.), 307-344, American Sciences Press, Inc., Columbus, Ohio.

Marsaglia, G. (1986). A current view of random numbr generators, Computer Science and Statistics: Proc. 16th Symp. Interface, Atlanta, March 1984 (Elsevier Science Publishers, Amsterdam).

Marsaglia, G., Zaman, A., and Tsang, W. W. (1990). Toward a universal random number generator. *Statistics and Probability Letters*, 8, 35-39.

Stein, C. M. (1945). A two-sample test for a linear hypothesis whose power is independent of the variance. {\em Annals of Mathematical Statistics}, 16, 243-58.

Tong, Y. L. (1980). *Probability Inequalities in Multivariate Distributions*. Academic Press, New York.

Wen, M. J. and Chen, H. J. (1994). Single-stage multiple comparison procedure under heteroscedasticity. *American Journal of Mathematical and Management Sciences*. Vol. 14, Nos 1 and 2, 1-48.